

BRIEF REPORT

Effect of Blinding and Unmasking on the Quality of Peer Review

Susan van Rooyen, BSc, Fiona Godlee, MRCP, Stephen Evans, MSc,
Richard Smith, FRCP, Nick Black, MD

The objectives of this study were to see whether, in the opinion of authors, blinding or unmasking or a combination of the two affects the quality of reviews and to compare authors' and editors' assessments. In a trial conducted in the *British Medical Journal*, 527 consecutive manuscripts were randomized into one of three groups, and each was sent to two reviewers, who were randomized to receive a blinded or an unblinded copy of the manuscript. Review quality was assessed by two editors and the corresponding author. There was no significant difference in assessment between groups or between editors and authors. Reviews recommending publication were scored more highly than those recommending rejection.

KEY WORDS: peer review; quality; biomedical; journal.
J GEN INTERN MED 1999;14:622-624.

Despite its central role, little research has been conducted into the relative effectiveness of different approaches to peer review.^{1,2} There are several reasons for believing that blinding may be beneficial: it may provide less biased reviews³; it may improve the quality of reviews, as judged by editors^{1,4-9} (a belief supported by one small randomized trial,¹⁰ but not by two larger ones^{11,12}); and papers published in journals that use blinded review are more likely to be cited.¹³ In turn, unmasking the identity of reviewers to one another may result in higher quality, though the one randomized trial to study this found no significant impact, as judged by editors.¹¹

This article describes a randomized trial to determine the effect of blinding, unmasking, and a combination of the two on review quality as assessed by the authors of the manuscripts. We also assessed the potential impact both of a Hawthorne effect and of the reviewers' recommendation regarding publication, and compared authors' ratings of review quality with those of editors.

Received from the *British Medical Journal*, London, U.K. (SVR, FG, RS), and London School of Hygiene and Tropical Medicine, London, U.K. (SE, NB).

Previous results from the same study were presented at the International Congress on Biomedical Peer Review and Global Communications, Prague, Czech Republic, September 1997.

Address correspondence and reprint requests to Susan van Rooyen: BMJ Editorial, BMA House, Tavistock Square, London WC1H 9JR, U.K.

METHODS

Consecutive manuscripts reporting original research received by the *British Medical Journal* and sent by editors for peer review between January and June 1997 were eligible for inclusion. Manuscripts were randomized into one of three groups: masked, unmasked, and uninformed. Each manuscript was sent to two reviewers who were selected by the editor responsible for the particular manuscript. In both the masked and unmasked groups, the reports of pairs of reviewers were exchanged. Reviewers in the unmasked group were asked to consent to their identity being revealed to their co-reviewer, whereas reviewers in the masked group remained anonymous to one another. The reviewers in the masked and unmasked groups were further randomized to receive either a blinded or an unblinded version of the manuscript. Blinding consisted of removing authors' details from the title page and acknowledgments. Blinded reviewers were asked whether they thought they knew the identity of the author or authors and, if so, to detail the names or institutions or both and to explain why they thought they could tell. As awareness of being in a study might affect the reviewers' behavior, the uninformed group was included, which allowed us to test for a Hawthorne effect. Manuscripts in the uninformed group were sent to two reviewers (masked and unblinded), who were not informed that a study was taking place.

On receipt of both reviews of a manuscript, the authors' details were removed from the reviews and their quality was independently assessed by two editors using a validated Review Quality Instrument.¹⁴ A decision on whether to publish the paper was made in the journal's usual manner. At least 10 days after decisions had been communicated to the authors, the corresponding authors were asked to evaluate the quality of the two reviews, using the Review Quality Instrument, and whether they thought each reviewer had been blinded to their identity.

The Review Quality Instrument (Version 3) consisted of seven items (importance of the research question, originality, method, presentation, constructiveness of comments, substantiation of comments, and interpretation of results), each scored on a 5-point Likert scale (1 = poor, 5 = excellent). A total score was based on the mean of the seven item scores. In addition, a global item seeking an assessment of the overall quality of the review was in-

cluded. The quality of each review was based on the means of the two editors' scores for each item and the total score and on the corresponding author's scores.

Analyses used independent comparisons of outcome measures between masked and unmasked reviewers, paired comparisons between blinded and unblinded reviewers, and independent comparisons between masked unblinded reviewers and uninformed reviewers, using *t*-tests. Paired comparisons were also made between authors' and editors' scores using *t* tests.

RESULTS

Recruitment and Response

Of an estimated 570 eligible manuscripts sent for peer review, 527 (92%) were entered into the study. The remaining 43 manuscripts were lost from the study, mainly as a result of administrative error. Of the 527 manuscripts randomized, 467 received two reviews: 149 masked, 160 unmasked, and 158 uninformed. Of the 160 in the unmasked group (320 reviewers), 10 did not have the reviewers' consent to their identity being revealed, and these manuscripts were not included in the analyses. The corresponding author provided review quality assessment of both reviews for 359 manuscripts (77%), and of only one review for 2 manuscripts, giving a total of 720 authors' assessments. Respondents were more likely to have their papers accepted for publication, subject to revision (50% vs 42%).

The characteristics of the papers (geographic origin) and the reviewers (as regards the factors known to be associated with review quality) were similar between groups.¹⁵ Exclusions did not appear to introduce bias.

Effect of Blinding and Unmasking on Authors' Opinion of Review Quality

Authors' assessments showed no statistically significant difference in the mean total score between masked

(*M*) and unmasked (*U*) groups ($M - U = -0.08$; 95% confidence interval [CI] $-0.23, 0.07$) or between blinded (*B*) and unblinded (*U*) groups ($B - U = -0.05$; 95% CI $-0.19, 0.09$) (Table 1). Similar results were found for individual items.

Analyses comparing successfully blinded reviewers (132 reviewers) with those who were randomized to be unblinded (245 reviewers) showed similar results (mean total score 2.82 vs 2.89; 95% CI $-0.21, 0.04$; $p = .07$).

Extent of a Hawthorne Effect

There was no significant difference between masked unblinded (*MB*) and uninformed (*U*) reviewers ($MB - U = -0.002$; 95% CI $-0.19, 0.19$), suggesting no important Hawthorne effect to take into account (Table 1).

Impact of Reviewers' Recommendation for Publication on Assessment of Review Quality

There was a statistically significant difference between authors' assessments of reviews that recommended publication (with or without revision) (*P*) and those recommending rejection (*R*) ($p < .01$) ($P - R = 0.39$; 95% CI $0.25, 0.54$), but no significant difference for editors ($P - R = -0.03$; 95% CI $0.14, 0.08$).

Comparison Between Editors' and Authors' Assessment of Review Quality

There was little difference in mean total scores (2.84 vs 2.89) (Table 2), and there were no editorially significant differences for individual items. Authors rated the reviewers' comments on the importance of the research question and the originality of the manuscript more highly than did editors. The reverse was true for the other items.

Table 1. Comparison of Review Quality for Blinded Versus Unblinded Reviewers, Masked Versus Unmasked Reviewers, and Masked Unblinded Versus Uninformed Reviewers in the Opinion of Authors (Range of Scores 1–5)

Item	Blinded, Mean (n = 245)	Unblinded, Mean (n = 245)	Difference	95% CI	Masked, Mean (n = 230)	Unmasked, Mean (n = 246)	Difference	95% CI	Masked Unblinded, Mean (n = 114)	Uninformed, Mean (n = 228)	Difference	95% CI
Importance	2.74	2.79	-0.05	(-0.26, 0.15)	2.78	2.76	0.02	(-0.21, 0.24)	2.69	2.76	-0.07	(-0.35, 0.22)
Originality	2.43	2.54	-0.11	(-0.32, 0.10)	2.39	2.59	-0.20	(-0.42, 0.02)	2.36	2.43	-0.06	(-0.34, 0.22)
Methodology	3.05	3.08	-0.02	(-0.21, 0.16)	2.96	3.15	-0.19	(-0.39, 0.003)	2.93	2.99	-0.06	(-0.31, 0.19)
Presentation	2.65	2.76	-0.11	(-0.32, 0.10)	2.65	2.73	-0.08	(-0.30, 0.14)	2.72	2.64	0.09	(-0.18, 0.36)
Constructiveness of comments	3.11	3.19	-0.08	(-0.30, 0.15)	3.13	3.18	-0.05	(-0.27, 0.17)	3.10	3.10	0.00	(-0.27, 0.27)
Substantiation of comments	2.91	2.99	-0.08	(-0.28, 0.12)	2.92	2.95	-0.03	(-0.26, 0.19)	2.92	2.86	0.06	(-0.22, 0.35)
Interpretation of results	2.98	2.89	0.09	(-0.11, 0.29)	2.91	2.93	-0.02	(-0.23, 0.19)	2.83	2.80	0.02	(-0.25, 0.29)
Mean total score (1–7)	2.84	2.89	-0.05	(-0.19, 0.09)	2.82	2.90	-0.08	(-0.23, 0.07)	2.79	2.80	-0.002	(-0.19, 0.19)
Overall quality	2.98	3.01	-0.03	(-0.24, 0.18)	2.97	3.02	-0.04	(-0.26, 0.17)	2.90	2.97	-0.07	(-0.34, 0.19)

Table 2. Comparison of Authors' and Editors' Assessments of Review Quality (Range of Scores 1–5)

Item	Authors (n = 720)		Editors (n = 720)		Difference	
	Mean	SD	Mean	SD	Mean	(95% CI)
Importance	2.76	1.24	2.54	0.93	0.22	(0.14, 0.31)
Originality	2.47	1.23	2.35	1.15	0.12	(0.03, 0.21)
Methodology	3.04	1.09	3.20	0.99	−0.16	(−0.25, −0.07)
Presentation	2.69	1.21	2.75	1.03	−0.06	(−0.16, 0.03)
Constructiveness of comments	3.13	1.21	3.30	0.85	−0.17	(−0.26, −0.07)
Substantiation of comments	2.92	1.23	3.04	0.96	−0.13	(−0.22, −0.03)
Interpretation of results	2.89	1.17	3.04	1.02	−0.15	(−0.24, −0.06)
Mean total score (1–7)	2.84	0.82	2.89	0.71	−0.05	(−0.11, 0.02)
Overall quality	2.98	1.18	3.25	0.88	−0.27	(−0.36, −0.18)

DISCUSSION

In the opinion of authors, there was little or no difference in review quality between masked and unmasked groups and between blinded and unblinded groups. There was no apparent Hawthorne effect and no significant difference in the mean total scores of editors and authors. Although statistically significant differences were found between authors' and editors' ratings for individual items, none was editorially significant. As might be expected, authors rated the reviewers' opinions as to the importance and originality of their papers higher than did editors. Authors' ratings of reviews that recommended publication were higher than those that suggested rejection.

The study has some methodologic limitations: it was undertaken in a large, frequently cited medical journal, and we cannot judge generalizability to other types of journals; although the response rate from authors was good, any respondent bias would have been expected to have increased the mean rating of reviews, so the difference between authors' and editors' ratings (mean total score difference = 0.05) might be slightly underestimated; the study is restricted to assessing the content of the review, not its impact on manuscript quality or the correctness of the opinions expressed; and at the time they made their assessment, authors were already aware of the fate of their paper.

Turning to the implications of our findings, blinding and unmasking appear to offer little or no benefit in this general medical journal as regards the quality of the reviews. Because authors derive no obvious benefit from reviewers being blinded or unmasked, there is no support from the results of this research for adopting either blinding or unmasking as policy. A decision to blind or unmask has, therefore, to be based on ethical considerations. This study suggests it is feasible for review quality to be assessed either by editors or by authors. Care should, however, be taken when comparing the results of studies that used editors' assessments with those based on authors' views. With the exception of originality and importance, the latter tend to rate reviews less favorably than is true of editors.

The authors thank all the BMJ authors and reviewers who so willingly participated; editors Tony Delamothé, Luisa Dillner, Trish Groves, Sandra Goldbeck-Wood, Tessa Richards, Roger Robinson, Jane Smith, Tony Smith, and Alison Tonks; and papers secretaries Sue Minns and Marita Batten.

Funding was received from the NHSE North Thames Research & Development Responsive Funding Group, London, U.K.

REFERENCES

- Lock S. A difficult balance: editorial peer review in medicine. London, UK: Nuffield Provincial Hospitals Trust; 1985.
- Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. *JAMA*. 1994;272:96–7.
- Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA*. 1994;272:143–6.
- Strasburger VC. Righting medical writing. *JAMA*. 1985;254:1789–90.
- Yankauer A. Peer review again. *Am J Public Health*. 1982;72:239–40.
- Shapiro S. The decision to publish: ethical dilemmas. *J Chronic Dis*. 1985;38:365–72.
- Ingelfinger FJ. Peer review in biomedical publication. *Am J Med*. 1974;56:686–92.
- Robin ED, Burke CM. Peer review in medical journals. *Chest*. 1987;91:252–5.
- Feinstein AR. Some ethical issues among editors, reviewers and readers. *J Chronic Dis*. 1986;39:491–3.
- McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. *JAMA*. 1990;263:1371–6.
- van Rooyen S, Godlee F, Evans S, Smith R, Black NA. The effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*. 1998;280:234–7.
- Justice AC, Cho MK, Winker M, et al. Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA*. 1998;280:240–2.
- Laband DN, Piette MJ. A citation analysis of the impact of blinded peer review. *JAMA*. 1994;272:147–9.
- van Rooyen S, Black N, Godlee F. Development of the Review Quality Instrument (RQI) for assessing peer reviews of manuscripts. *J Clin Epidemiol*. 1999;52:625–9.
- Black N, van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and what makes a good review in a general medical journal? *JAMA*. 1998;280:231–3.